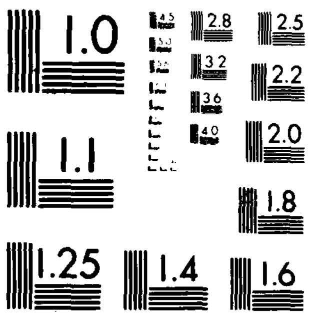
AD-A094 275 PENNSYLVANIA STATE UNIV UNIVERSITY PARK APPLIED RESE--ETC F/G 17/2
DIGITAL PROCESSING OF SPEECH MATERIALS: A CRITICAL-BAND-BASED M--ETC(U)
JUL 80 R D CELMER N00024-79-C-6043
UNCLASSIFIED ARL/PSU/TM-80-180 NL

END
DATE
FILMED
2 81
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A

6

DIGITAL PROCESSING OF SPEECH MATERIALS: A CRITICAL-BAND-BASED MODEL OF SPEECH PERCEPTION

10 Robert D./Celmer

9 Master's thesis

Technical Memorandum
File No. TM 80-180
11 Jul[illegible]80
Contract No. N00024-79-C-6043

Copy No. 5 15

12 PGI

14 ARL/PSU/TM-80-180

The Pennsylvania State University
Intercollege Research Programs and Facilities
APPLIED RESEARCH LABORATORY
Post Office Box 30
State College, PA 16801

NAVY DEPARTMENT

NAVAL SEA SYSTEMS COMMAND

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TM 80-180	2. GOVT ACCESSION NO. AD-A094 275	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DIGITAL PROCESSING OF SPEECH MATERIALS: A CRITICAL-BAND-BASED MODEL OF SPEECH PERCEPTION		5. TYPE OF REPORT & PERIOD COVERED MS Thesis, November 1980 6. PERFORMING ORG. REPORT NUMBER TM 80-180
7. AUTHOR(s) Robert D. Celmer		8. CONTRACT OR GRANT NUMBER(s) N00024-79-C-6043
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Pennsylvania State University Applied Research Laboratory P. O. Box 30, State College, PA 16801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Sea Systems Command Department of the Navy Washington, DC 20362		12. REPORT DATE July 16, 1980 13. NUMBER OF PAGES 59 pages & figures
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified, Unlimited 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited, per NSSC (Naval Sea Systems Command), 9/23/80

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

thesis, speech, hearing, perception

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Existing literature suggests that the hearing mechanism deals with incoming speech material by filtering the signals into a series of frequency bands. The width of these bands has been referred to as the critical band that is the perceptual frequency bandwidth observed in a variety of psychoacoustic contexts. Digital processing techniques have been developed for altering available recorded speech materials so that the frequency resolution available in the resultant stimuli may be

20. ABSTRACT (continued)

controlled. Tapes have been produced wherein the frequency bandwidth resolution is limited to no better than one critical band and these tapes have been used in intelligibility testing. Some existing research indicates that the critical band is significantly widened in many individuals with sensorineural hearing loss of cochlear etiology. The digital processing routines described above were also used in developing tape recorded materials with bandwidth resolution limits considerably wider than the normal critical band. The bandwidths chosen for this stage of the digital processing were based on empirical observations of the critical band of sensorineural hearing impaired patients. These recordings were also used in intelligibility testing with normal listeners. Implications of these studies for the clinical measurement of speech intelligibility will be discussed.

ABSTRACT

Existing literature suggests that the hearing mechanism deals with incoming speech material by filtering the signals into a series of frequency bands. The width of these bands has been referred to as the critical band that is the perceptual frequency bandwidth observed in a variety of psychoacoustic contexts. Digital processing techniques have been developed for altering available recorded speech materials so that the frequency resolution available in the resultant stimuli may be controlled. Tapes have been produced wherein the frequency bandwidth resolution is limited to no better than one critical band and these tapes have been used in intelligibility testing. Some existing research indicates that the critical band is significantly widened in many individuals with sensorineural hearing loss of cochlear etiology. The digital processing routines described above were also used in developing tape recorded materials with bandwidth resolution limits considerably wider than the normal critical band. The bandwidths chosen for this stage of the digital processing were based on empirical observations of the critical band of sensorineural hearing impaired patients. These recordings were also used in intelligibility testing with normal listeners. Implications of these studies for the clinical measurement of speech intelligibility will be discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

The author wishes to extend a gracious "thank you" to all those who gave so much of their time and effort to help make this endeavor a success, especially,

Dr. Gordon Bienvenue, from whom I have come to better understand the meaning of such words as "mentor," "apprenticeship," "inspiration," and "fellowship;"

Remingius Onyshczak, whose approach to assistantship is a joy to experience;

Virginia Lolla, whose ilse of view constantly resurged feelings of self-confidence;

Dr. Walter Celmer, whose encouragement and love helped the author to initially pursue graduate studies, as well as

Bryan Jensen, Robert Hirlinger, James Prout, Daniel Richards, Dr. Paul Michael, Fran Murphy, the 1979-1980 class of SPA graduate students and the Spirit within all things.

This research was supported by the Applied Research Laboratory at The Pennsylvania State University under contract with the Naval Sea Systems Command.

This document is dedicated to the memory of my mother, Florence Mack Celmer.

Chapter I

BACKGROUND

Brief Physiology Overview

Detailed descriptions of auditory anatomy may be found in such sources as Fex (1962), Goldstein (1968), Milner (1970), Dallos (1973) and Spoendlin (1973). The portion of the inner ear primarily responsible for the hearing process is called the cochlea. This coiled, membraneous structure contains three compartments, or scalae, separated lengthwise along the human cochlea's two-and-three-quarter turns. The scala media houses the Organ of Corti, whose complex set of receptor cells serve as the locality for conversion of continuous acoustic waves into neural impulses. An acoustic input arriving at the oval window via the ossicles gives rise to transverse and longitudinal waves in the cochlea's endo- and perilymphatic fluids, respectively. Wave propagation in the cochlea is essentially a case of wave propagation in a shallow fluid of nonuniform depth (i.e., similar to ocean waves approaching a beach). The scala tympani and the scala vestibuli become gradually "shallower" towards the apical end. The point at which the transverse wave at the fluid interface of these two scala (i.e. scala media duct) will crest is dependent on its frequency (Dallos, 1973). High frequency waves

crest near the basal end (far from "shore") while low frequency waves consequently crest close to the apical end. Maximal dissipation of the waves' energy occurs at this location. The inner and outer hair cells of the Organ of Corti are embedded in the basilar membrane. These hair cells have cilia (hairlike projections) which are rooted in a surface plate and the ends of these cilia either float freely in the endolymph or course into the opposite tectorial membrane. Wave motion causes the cilia to undergo shear, inducing the corresponding hair cell to trigger the neuronal fibers synapsed at its base. It should be noted that the hair cells existing at the wave's crest point will undergo maximal shear and thus relay the most neural information. Thus, the hydrodynamic construction of the cochlea combined with the morphology of the Organ of Corti yields a preliminary place-specific frequency analysis of the acoustic signal.

Bipolar cells synapse at the base of the hairs with neurons that carry frequency, phase, and amplitude information towards the brain stem. The collection of the cell bodies of these neurons is called the spiral ganglion, located in the modiolus. The subsequent collection of the axons proceeding to the brainstem make up the bulk of the auditory nerve (VIII). As the information travels to the cortex, it is relayed to the superior olives by the dorsal and ventral cochlear nucleii. From superior olives information is fed to the inferior colliculi via the lemniscal pathways. Information is then transmitted via the brachium of the inferior colliculus to the medial geniculate bodies of the thalmus and then via auditory radiations to the temporal cortex. Decussations take place at the level of the

dorsal cochlear nucleus, the superior olives, and the inferior colliculus. The first stop for an afferent message arriving at cortex is Broadman areas 41 and 42 (auditory area A-I of Woolsey and Walzl, 1942) of the temporal lobe, located anterior to the calcarine fissure.

The existence of an efferent auditory pathway was discovered near the turn of the century by Ramon Y Cajal (1896) and elucidated by Lorente de No (1937). Originating in the posterior cephalad of the diencephelon, the downward coursing fibers experience much the same afferent relays in the opposite order (Fex, 1962). Rasmussen (1946) made a highly detailed account of that portion of the efferent pathway progressing from the superior olives to the cochlea, naming it the olivo-cochlear bundle. The most peripheral part of this tract branches and spirals to distribute itself peripherally to all turns of the Organ of Corti. Subsequent research by Fernandez (1951) noted that the branched fibers innervate a diffuse group of inner and outer hair cells.

The Critical Band Phenomenon

Existing literature on audition theory suggests a heavy reliance upon the notion of critical bands (e.g., Scharf, 1970). Most simply, a critical band may be conceived as an internal bandpass filter. In the initial conception of critical bands (Fletcher, 1940), the auditory system was theorized as consisting of a fixed bank of about twenty-four critical bands laid end to end, covering the audible frequency range. By contrast, current notions view critical

bands as variable filter elements centered upon a particular signal frequency (cf. Scharf, 1970). The critical band has been defined empirically as "that bandwidth at which subjective responses rather abruptly change" (cf. Scharf, 1970, p. 159). In general, two stimuli separated in frequency by less than a critical bandwidth will interact in one of a number of ways, while two stimuli separated by more than a critical bandwidth will not. Changes of listener response due to the critical band phenomenon have been observed in such perceptual phenomena as masking, loudness, and musical consonance.

Masking and the Critical Band. A condition of masking is said to be in effect when a temporary loss of sensitivity to a stimulus occurs, caused by a simultaneous, ipsilateral presentation of another stimulus (Moore, 1977). By and large, a masking stimulus is most effective in hiding a given signal whenever the frequency content of the two signals is similar (Scharf, 1970). Under this context, then, a critical band may be defined as that masker frequency region wherein the masking of a given stimulus is most effective (Scharf, 1970). Energy concentrations in regions larger than the critical band of the test signal demonstrate less efficient masking ability. A masker's efficiency may be defined as that amount of masking energy needed to provide a given threshold of masking. In a psychoacoustic context, a masking source is most efficient when its bandwidth lies within the critical band. Thus, the critical bandwidth mechanism seems to provide a resolution capability to the auditory system whose limit is reflected in its ability to mask signals.

<u>Loudness and the Critical Band</u>. Concomitant to the limits of masking bandwidth is the way a listener perceives the intensity of noise bands as a function of frequency bandwidth (Δf). Studies by Zwicker and Feldtkeller (1955) and Zwicker (1958) showed a significant relationship between the bandwidth of the stimulus and the point at which a listener hears a change in loudness:

> . . . the loudness of a subcritical complex sound of invariant intensity is largely independent of Δf-- it is about as loud as an equally intense pure tone lying at the band's center frequency. Only when Δf exceeds the critical band does the loudness of the complex begin to increase (Scharf, 1970).

The process of integrating incoming sound stimuli for loudness perception, then, appears to operate through the filtering effect of the critical bandwidth mechanism.

<u>Musical Consonance</u>. Esthetic tests which rate the pleasantness of two-tone complexes have provided another setting under which the critical band phenomenon may be observed. Listeners were asked to rate the consonance of a pair of tones on a seven-point "pleasantness" scale. The overriding judgments of consonance were found at tonal separations of more than a critical band. A rapid decrease in the rating occurred as the tones moved to a separation narrower than a critical band (Plomp, 1964; Greenwood, 1961). These findings psychoacoustically demonstrate a phenomenon musicians have understood for centuries; consonance peaks occur at the common harmonic ratios--the fourth, fifth and octave (Plomp and Levelt, 1965). The relationship of these ratios to the critical band phenomenon in particular further demonstrates the universality of this effect in

hearing.

Critical Bands and Speech. Speech is the most pervasive and important acoustic stimulus for the human listener. Evidence suggests that the critical band may serve in the analysis of speech (cf. Scharf, 1970). The specific task of attempting to measure the critical bandwidth used in the analysis of speech sounds by the auditory system has been indirectly approached in the work of French and Steinberg (1947). The twenty-four critical bands found in pure tone psychoacoustic studies were found to match very closely to the twenty-four bands which contributed equally to speech intelligibility. In summarizing other research on critical bands, one finds two functional aspects of critical bands which appear to play an important role in the perception of speech and hence, represent fundamental aspects of a listener's performance. As implied in the works of Fletcher (1940), Zwicker (1958), and Greenwood (1961), the critical band serves to band limit background noise. The narrower the pass-band of the ear as a filter, the more noise the ear can reject, making it more tolerant to lower signal-to-noise ratios. Thus, a listener may be able to correctly perceive a spoken communication despite background noise, simply because much of the energy associated with the noise lies outside the critical bands surrounding the formant frequencies of the speech.

Secondly, our ability to discriminate the harmonic content of complex signals (one of the many cues used, for instance, in speaker identification) is similarly related directly to the critical band phenomenon. Plomp (1964) has demonstrated that listeners are

able to discriminate those partials of a complex tone which lie more than a critical bandwidth apart. Morton and Carpenter (1963) found that formants can be identified by listeners even when no prominent energy peak is present as long as the most intense harmonics associated with each formant are separated by at least a critical bandwidth. Synthetic vowels presented to listeners by Remez (1977) showed an abrupt changeover from speech-like to non-speech-like sounds as the formant bandwidth increased to greater than a critical bandwidth. Preliminary analysis with reference to the critical bandwidth phenomenon indicates that this mechanism seems to be working for speech analysis on a peripheral basis.

Two functional characteristics of critical bands have been briefly reviewed (noise band limiting and harmonic discrimination). It is argued from these effects that critical bands play an important role in the correct perception of complex acoustic stimuli, such as speech.

A Mechanism for Critical Bands. A selective inhibition of frequency-specific afferent auditory messages by the efferent olivo-cochlear bundle may serve as the neuronal basis for the critical band phenomenon. This gating effect by neuronal inhibition has been clearly demonstrated to exist (cf. Rasmussen, 1946; Desmedt and Monaco, 1961; Fex, 1962). This suppression of neural response may occur at:

1. individual afferent fibers of inner hair cells (post-synaptic), and

2. the entire output of individual outer hair cells (pre-synaptic) (Spoendlin, 1977).

Fex (1967) proposed a network describing the afferent and efferent pathways involved with each cochlea. From this morphology, a hypothetical feedback mechanism may be proposed (see Figure 1). The ascending pathway courses from the cochlea's hair cells through the VIII nerve to the cochlear nucleus and then to the superior olives via the olivo-cochlear bundle to the cochlea. Since there is a 33:1 ratio of afferent to efferent fibers, there is a limit to which frequency information may be gated. Scharf (1970) reported twenty-four critical bands (narrow frequency regions of gated information) whose widths equal approximately 16 percent of center frequency for those bands above 500 Hz (see Table 1).

It would follow that any decrement in the efficiency of this feedback mechanism would incur deficits in the functional characteristics of the critical band mechanism described above. Several studies (Fex, 1967; Dewson, 1968; Capps and Ades, 1968; Trachiotis and Elliott, 1970; Pickles, 1973) have examined the behavioral performance of animals in which the action of olivo-cochlear bundle was blocked. Two overriding observations were made:

1. reduction in frequency discrimination ability, and
2. reduction in the ability to recognize signals in a background noise.

Thus, it is reasonable to assume that units of the neural inhibitory system of the cochlea would be damaged whenever significant cochlear pathology is present. In such a case, one would expect to find

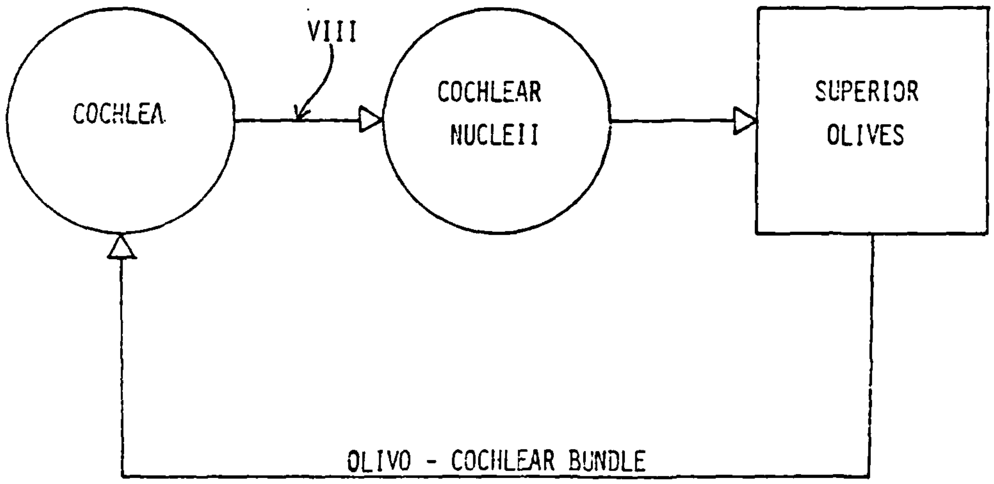


FIGURE 1. HYPOTHETICAL FEEDBACK MECHANISM.

Table 1

Examples of Critical Bandwidth

Number	Center Frequency (Hz)	Critical Band (Hz)	Lower Cutoff Frequency (Hz)	Upper Cutoff Frequency (Hz)
1	50	--	--	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1,000	160	920	1,080
10	1,170	190	1,080	1,270
11	1,370	210	1,270	1,480
12	1,600	240	1,480	1,720
13	1,850	280	1,720	2,000
14	2,150	320	2,000	2,320
15	2,500	380	2,320	2,700
16	2,900	450	2,700	3,150
17	3,400	550	3,150	3,700
18	4,000	700	3,700	4,400
19	4,800	900	4,400	5,300
20	5,800	1,100	5,300	6,400
21	7,000	1,130	6,400	7,700
22	8,500	1,800	7,700	9,500
23	10,500	2,500	9,500	12,000
24	13,500	3,500	12,000	15,500

wider-than-normal critical bands due to the loss of inhibitory units.

There is no reason to assume that the central nervous system is aware of the particular details of a peripheral pathology. Thus, the central nervous system expects to receive information from a full complement of critical bands. Hence, the widening of individual critical bands involves two factors:

1. a broadening of the integration region which results in each critical band integrating a larger area for a given signal;

2. retention of the complete number of critical bands such that the critical bands may be expected to overlap one another in the widened case.

Therefore, the energy content of frequency regions common to more than one band will be integrated more than once. This phenomenon may be expected to lead to an abnormally high perception of loudness for a given magnitude of input acoustic energy. This phenomenon is observed psychoacoustically among sensorineural hearing impaired individuals, and is referred to as recruitment (cf. Fowler, 1928; Michael and Bienvenue, 1976).

Bonding (1979) observed indications of widened critical bands in some 50-67 percent of the sensorineural hearing impaired listeners which he examined. In addition, Bonding's data demonstrate that the width of the widened critical band is independent of the magnitude of threshold hearing loss amongst those sensorineurals with critical bandwidth distortion. This finding has been supported in tests by Michael and Bienvenue (1976); Bienvenue and Michael (1977); and Bennett et al., (1978), who found evidence of

widened critical bands in noise exposed patients which was not correlated to threshold shift magnitudes. In fact, some critical bandwidth distortion occurs in the absence of threshold shift (Michael and Bienvenue, 1976).

The two functional characteristics of this mechanism suggest the symptoms that may appear in a cochlear pathology. A common finding among individuals with cochlear hearing loss is that relatively small amplitudes and remote frequencies of background noise are detrimental to speech perception. This phenomenon may very well follow directly from the reduced band limiting capabilities of the widened critical bands (Michael and Bienvenue, 1976). In addition, it is clear that such a pathology, resulting in a widening of critical bands, will tend to reduce the number of discriminable harmonics of a signal such as speech. Listeners with this problems will be less able to discriminate speech on the basis of its harmonic content. Subjects with cochlear pathologies report speech to sound "foggy" or "blurred," with some insisting that everyone mumbles when they speak (cf. Fowler, 1928; 1937). The integrity of the listener's critical bands, therefore, appears to represent a limiting factor in their ability to perform these everyday tasks.

The plight of this pathology lies in its lack of remedial measures. While this phenomenon has existed amongst the general populace for as long as the classic case of threshold loss, no present-day audiometric aids are available that can alleviate these symptoms. Simply amplifying the signal to try to compensate for such deficiencies only worsens the effect by presenting both the

desired speech and the unwanted background noise equally loud. Rather, a means to test for early signs of bandwidth widening might prevent extensive damage and provide a pool of knowledge on which to base remedial measures. Standard laboratory procedures for measuring critical bandwidth typically involve lengthy psychophysical techniques performed on trained subjects (e.g., Fletcher, 1940; Zwicker, 1954; Greenwood, 1961; Haggard, 1974). These procedures, while extremely powerful and accurate, are not easily applied to the clinical setting where listeners are untrained and unwilling to spend the requisite time listening to sophisticated signal complexes. Development of a clinical testing procedure which determines critical bandwidth directly and rapidly should prove more desirable than the standard tests.

Chapter II

STATEMENT OF THE PROBLEM

The condition of widened critical bands should present a disruption to the source-path-receiver communication chain. Given a clearly articulated speech signal propagating through a conducive acoustic medium, one finds a receiver whose frequency "resolving power" is insufficient to enable discrimination of that speech from the entire acoustic stimulus. The degree to which the listener has lost his/her resolving power is a variable (N = factor by which the critical band is widened, see Figure 2a). The approach chosen to determine N involves adding an interruptive processor to the communication chain, whose resolving characteristics are known and can be varied at will (see Figure 2b). By observing the response of the unknown receiver while varying the known disruption of the presentation, N may be determined. If the frequency resolution of the interruptive processor is finer than the resolving power of the listener, then perception is limited simply by the listener's critical bandwidth. Widening the bandwidth of frequency resolution for the processor should have no effect upon the listener's acoustic analysis of the signal (compared to the unprocessed case) until the resolution becomes coarser than the listener's critical bandwidth

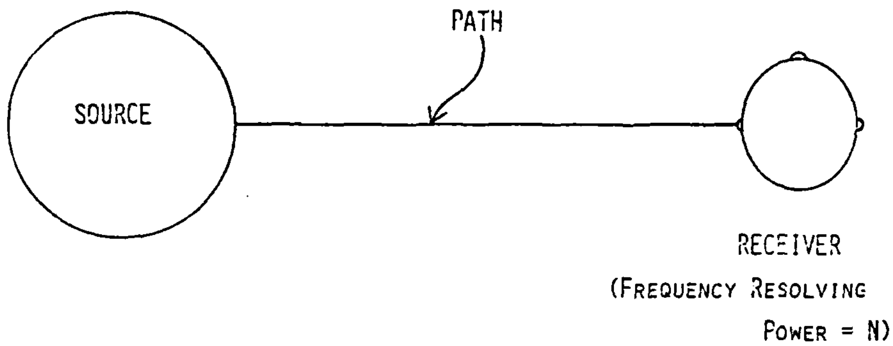


FIGURE 2A. COMMUNICATION CHAIN.

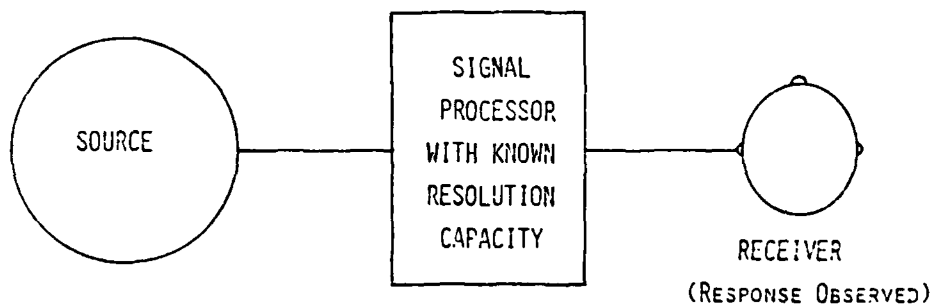


FIGURE 2B. COMMUNICATION CHAIN WITH INTERRUPTIVE PROCESSOR.

(i.e., his/her frequency resolution capacity), at which point a decrement in performance should be observed. Thus, the critical bandwidth of a listener in this procedure equals N-times the normal bandwidth capability, indicated by that resolution at which he/she first exhibits a decrement in performance.

The purpose of this study, then, is to:

1. Generate acoustic stimuli of varying resolutions equal to or greater than the normal critical bandwidth using digital signal processing; specifically,

    a. Is it feasible to produce bandwidth limited signals such as those described in Chapter III?

    b. Is it feasible to produce a processing scheme which allows for variation of bandwidths from the normal critical band to integer multiples of the critical band?

    c. Do these varying bandwidth limited speech signals demonstrate characteristics observed in speech processing by humans under pathologic conditions such as the phenomenon of recruitment?

2. Using these signals as stimuli for speech discrimination testing, specifically,

    a. Do normal listeners, presented with wider than normal bandwidth resolution limited signals, demonstrate performance decrements comparable to those seen in sensorineural hearing impaired listeners? That is, does a widened bandwidth condition effectively model sensorineural hearing impaired speech listening?

    b. Do normal hearing listeners demonstrate a monotonic trend of decreasing performance in speech intelligibility as their allowed bandwidth resolution is systematically widened? That is, does the <u>magnitude</u> of bandwidth widening effectively model the <u>magnitude</u> of impairment in speech discrimination?

Chapter III

DIGITAL SIGNAL PROCESSING OF SPEECH MATERIALS

General Overview

The generation of bandwidth resolution limited speech involves the digital signal processing algorithm shown in Figure 3. Prerecorded stimuli are input to a computer, processed, and then rerecorded onto audio tape in the processed form. The first step of this procedure requires the transformation of a continuous input into a series of discrete elements. The input has a continuously varying amplitude and is called an analog signal. The digital signal contains an array of discrete values corresponding to the input amplitude. The device used to transform the signal from a continuous to a digital mode is called an analog to digital (or A/D) converter. Within this converter is a timing pulse which beats at a fixed rate known as the sampling rate. As the analog signal is fed into the A/D converter via a conventional playback machine, the amplitude of the signal is detected at every occurrence of the timing pulse. The digital portion of a computer may receive and store these discrete amplitudes of the wave in the form of a one-dimensional array of voltages. For example, a digitized sine wave might have an array "A" equal to (0, 2, 4, 6, 4, 2, 0, -2, -4,

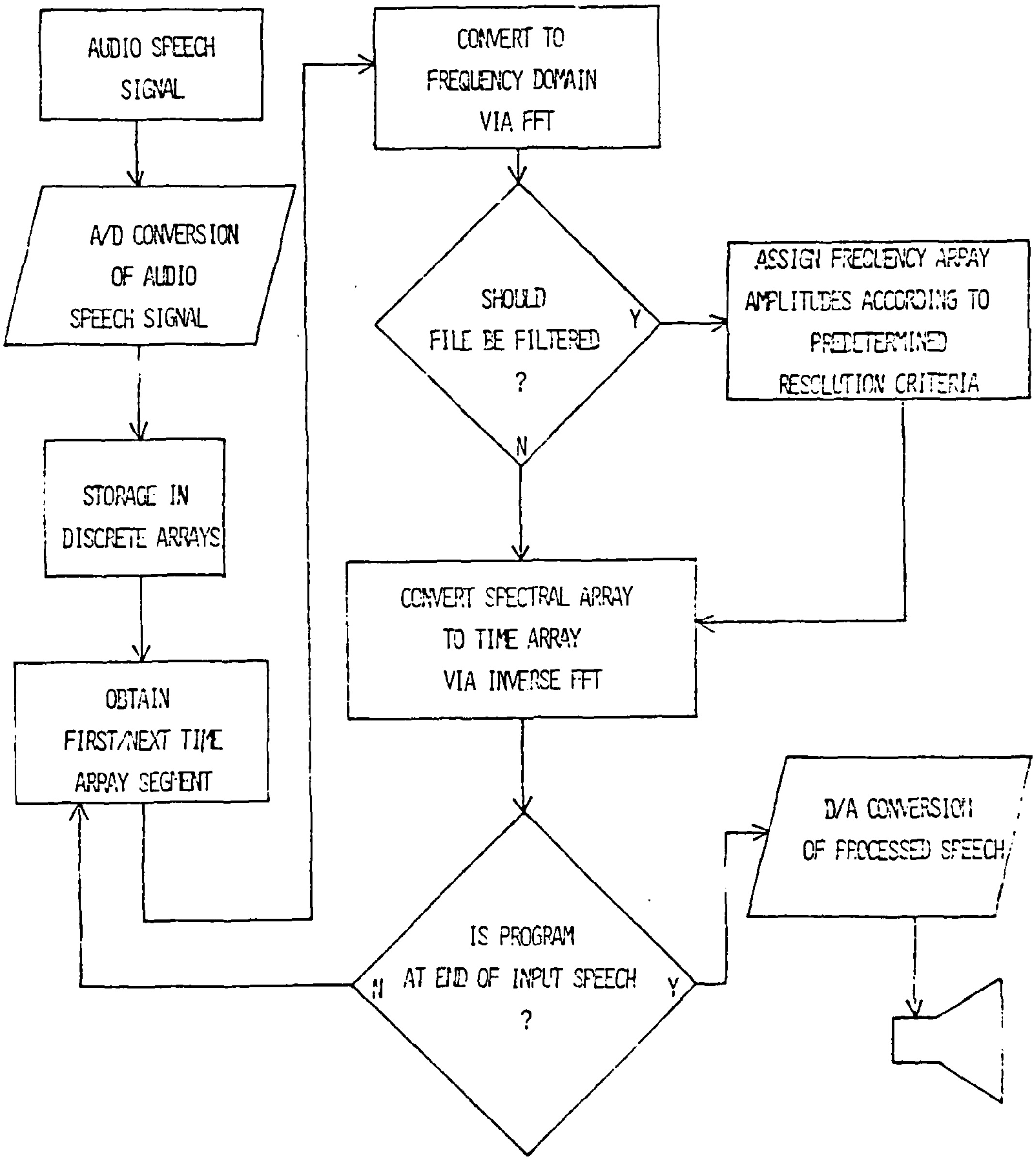


FIGURE 3. PROCESSING ALGORITHM.

-6, -4, ...). The sampling rate commonly used for audio signals is in excess of 20,000 samples per second. Thus, the example given above would be a typical array representing a pure tone at or above 2000 Hz (depending upon the precise sampling rate).

Once the word list is digitized and stored as a time domain array within a digital computer, the processing scheme may be initiated by the software program. The data are then converted in small time increments into the frequency domain by what is known as a Fast Fourier Transform.

The Fast Fourier Transform

A convenient and precise method for analyzing audio signals involves delineation by sums of sinusoids or complex exponentials. Commonly called Fourier representations, they provide an inherently superior tool to signal processing for two fundamental reasons. First, a linear system's response may be easily determined from a superposition of sinusoids or complex exponentials. Secondly, the Fourier representation often reveals properties of a signal that would otherwise be less evident (Rabiner and Schafer, 1978).

Early models for speech production of steady state vowels or fricatives, for example, all involved a linear system excited by either a periodic or random source. Consequently, Fourier analysis was utilized traditionally in the evaluation of such spectra. More recently, however, speech has been viewed as a much more dynamically complicated waveform (Ladefoged, 1962; Curtis, 1968; Minifie et al, 1973). The combined transient,

random and periodic nature of a speech signal induces marked changes in amplitude with time, violating the steady-state requirements of a standard Fourier representation. Instead, a short-time analysis principle applied to the Fourier method has been found to be a valid approach to speech processing (Rabiner and Schafer, 1978). These authors found that a steady state assumption for the spectral properties of speech is valid for time intervals on the order of 10-30 msec. In a review of this time-varying technique, the application of its principles to fast computation algorithms for discrete Fourier analysis (FFT algorithms) will be demonstrated.

Classical Fourier analysis of spectra has two basic approaches. For purely periodic waveforms, one determines its Fourier Series [see Equation 1]:

$$X(t) = a_o + \sum_{n=1}^{\infty}[a_m \cos(m\omega_o t) + b_m \sin(m\omega_o t)] \quad , \tag{1}$$

where: t = time,

T = the period of $X(t)$,

$\omega_o = 2\pi/T$,

$$a_o = 1/T \int_0^T X(t)dt \quad ,$$

$$a_m = 2/T \int_0^T X(t)\cos(m\omega_o t)dt \quad ,$$

and

$$b_m = 2/T \int_0^T X(t)\sin(m\omega_o t)dt \quad .$$

By contrast, pulse-like waveforms are analyzed by evaluation of its Fourier Transform [see Equation 2].

$$x(f) = \int_{-\infty}^{\infty} X(f)\exp(-j2\pi ft)dt \quad \text{or} \quad X(t) = \int_{-\infty}^{\infty} x(t)\exp(-j2\pi ft)df \ . \qquad (2)$$

Underlying the Fourier Series method is the following notion: the elemental periodic waveform is a sinusoid of the form:

$$x(t) = A\cos(2\pi ft - \theta).$$

Further, all periodic waveforms are consequently comprised of some unique summation of sinusoids. Each of the summation terms exists at the discrete frequencies given by $m\omega_o$. Each term is also harmonically related to the fundamental, $1/T$, by the index m. The two parameters that define that set are the amplitude spectrum and the phase spectrum. Whenever these two parameters are evaluated, whether electrically, mechanically, or mathematically, the process said to be occurring is called spectral analysis.

The Fourier Transform, on the other hand, defines x(f) as a continuous function of frequency. There is no index, m, as in the Fourier Series, which would have indicated a dependence upon discrete frequencies. Aperiodic, or pulse-like waveforms must have their complete time history integrated to determine the corresponding frequency composition.

While analyzing speech, however, one finds a mixture of both periodic and aperiodic waveforms (Minifie et al., 1973); neither method alone is complete. Rather, a Discrete Fourier Transform (DFT) capitalizes on the discrete nature of the waveform's amplitude to

enable provision of spectral information regardless of periodicity [see Equation 3]:

$$X_d(k\Delta f) = (1/M\Delta t)\sum_{m=0}^{M-1} x(m\Delta t)\exp(-j2km/M) \ . \qquad (3)$$

where: t = time,

T = time interval of sampling,

Δt = time sampling spacing,

Δf = frequency sampling interval = $1/T = 1/Mt$,

M = number of samples in T,

m = time index (0, 1, 2, ..., M-1) ,

and

k = frequency index (0, 1, 2,..., M-1) ,

If the source function $x(m\Delta t)$ repeats itself with time, the evaluation occurs as it would in a Fourier Series computation. In the case of a transient function, the array of distinct amplitude values capacitates a direct summation of the complex Fourier Transform. It should be noted that computation algorithms involving a series summation are much more efficiently realized by a computer than are formal evaluations of integrals. Thus the DFT, which serves as the basic algorithm of a Fast Fourier Transform (FFT), efficiently performs spectral analysis of speech signals, given adherence to certain necessary criteria, described below.

The main requirement for using the DFT is that the digitized speech waveform must satisfy the Nyquist sampling criterion; the sampling rate should be at least twice as great as the highest frequency in the waveform sampled (Rabiner and Schafer, 1978).

Sampling at twice the highest frequency contained in the input provides greater than two samples for each fundamental waveform; this insures proper coding of the signal's frequency. Violation of this rule leads to the phenomenon of aliasing, in which high frequency amplitude values are confused as low frequency information (see Figure 4). Foldover is a term which describes the magnitude of frequency displacement error induced by the aliasing phenomenon. That is, half the sampling frequency serves as a pivot frequency for aliasing in that the low frequency alias occurs at a frequency as far below the pivot frequency as the high frequency component is above the pivot frequency. For example, if a sampling rate of 20 kHz is used for speech (yielding a pivot frequency of 10 kHz), any high frequency component at 15 kHz is "folded down" to become a 5000 Hz low frequency alias, yielding an inaccurate spectrum (see Figure 5).

Another important consideration in the computation of the DFT within a software program is the time necessary to complete it. Cooley and Tukey (1965) found a significant reduction in the number of complex additions and multiplications needed for this transform whenever the number of samples chosen for each computation equaled a power of 2 (i.e., when $M = 2^n$). In addition, the two indices involved in the compilation always have a value of 0 or 1 for $M = 2^n$, a feature exploited by certain FFT subroutines to gain additional time advantage (Singleton, 1969).

An elementary property of the Discrete Fourier Transform is that it is a linear operation (Cooley et al., 1969). Stated

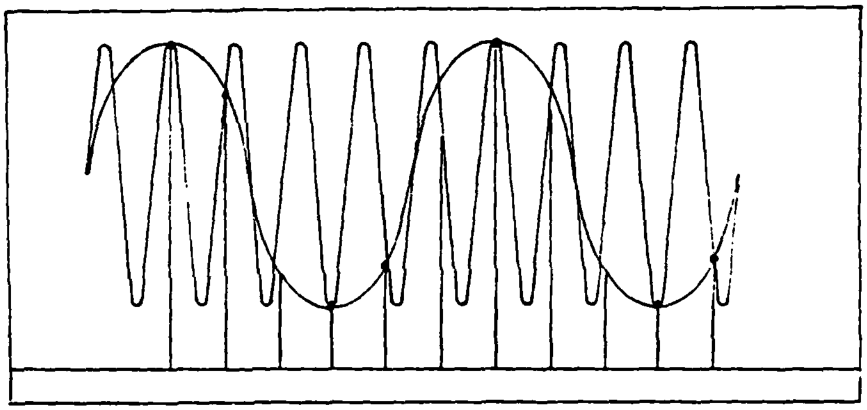


FIGURE 4. ALIASING PHENOMENON.

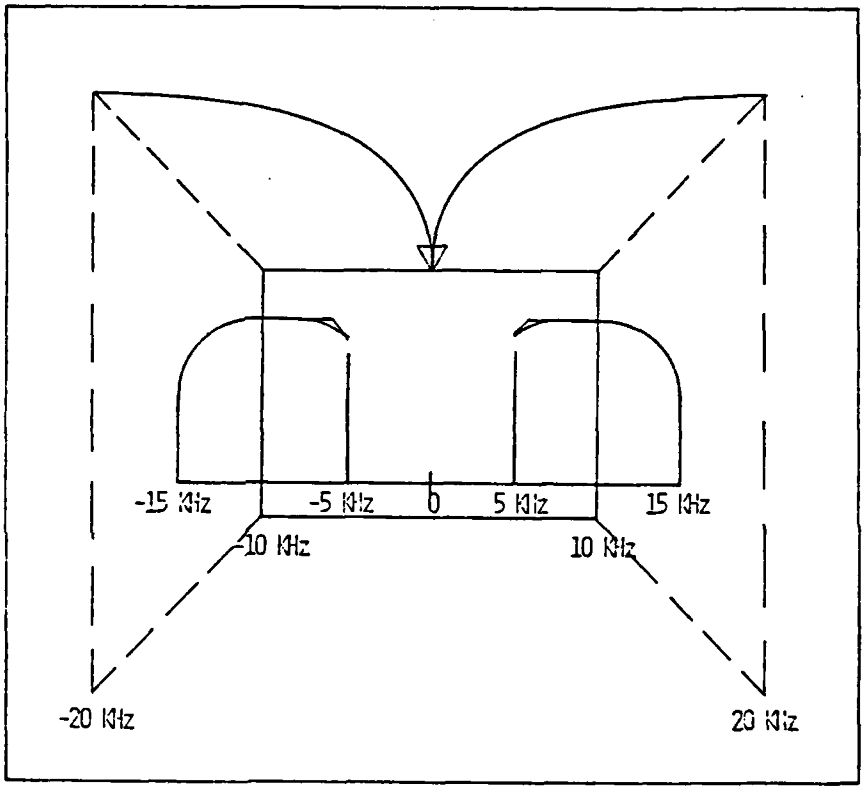


FIGURE 5. FOLDBACK PHENOMENON.

mathematically, $X_d(k\Delta f) \leftrightarrow x(m\Delta t)$, it indicates the validity in performing an inverse DFT. Since this research manipulates spectra while in the frequency domain, a viable method to gain access to and from that realm would be a necessary and sufficient requirement. Thus, the linearity of the DFT provides the symmetrical tool upon which such processing as digital filtering depends.

The FFT is the functional software realization of the Discrete Fourier Transform. Used as a subroutine, it makes available (in a forward transform) arrays corresponding to the real and imaginary componenets of a spectrum's amplitudes. Conversion of these values to polar form yields one magnitude for each frequency array element. There is a fixed frequency interval between the source values for each array element; for example, array element number one might correspond to the amplitude of that instantaneous signal at 70 Hz, while array element number two might correspond to the instantaneous 140 Hz amplitude, etc. The number of frequency elements depends on the sampling rate used to initially digitize the waveform, and the number of samples (time segment size) used in the FFT process. Sampling rates are generally in excess of 20,000 samples/second in order to satisfy the Nyquist criterion for the primary audio range (i.e., below 10,000 Hz), while time segments on the order of 10-30 msec are taken sequentially to approximate the steady state condition described earlier (cf. Rabiner and Schafer, 1978).

Note that this information is stored independent of specific frequency data. The affiliation of a voltage value with a

particular frequency is an arbitrary component of the output process and this stored array of voltages is independent of frequency information prior to output processing. Thus, the term "filtering" takes on a new meaning in the digital mode. Instead of running the signal through a relatively coarse analog filter, each instantaneous spectral array may be modified by simply specifying the energy content between predetermined frequency limits (see Figure 6). The slopes on digital "filters" are nearly infinite and permit the generation of tightly tuned, nonoverlapping band-pass filtering assignments like those found in the normal auditory periphery (cf. Scharf, 1970).

The processed spectra were made using the frequency limits recommended by Scharf (1970) and reported above in Table 1. The discrete frequency amplitudes that fall within each bandwidth are averaged; each of the discrete amplitudes of that band are then set equal to this r.m.s. value, limiting the resolution allowed to the preselected bandwidth for that time segment (see Figure 6). The bandwidths shown in Table 1 give the limits for a normal critical bandwidth (i.e., CB = 1X). Coarser filtering schemes are realized by multiplying the bandwidths by a chosen integer value (retaining the original center frequency), and averaging the amplitudes contained within these widened limits.

Once the frequency assignments have been made, the new array of spectral magnitudes are converted to rectangular form and placed in a call to an inverse FFT.

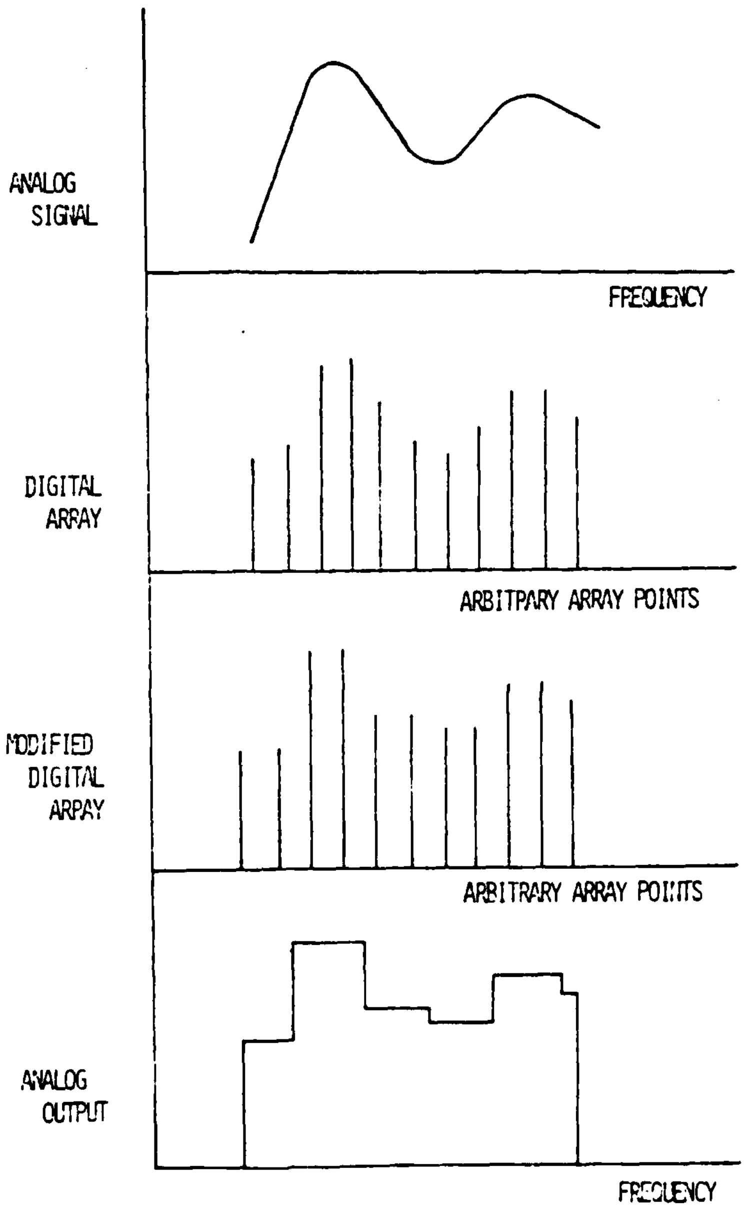


FIGURE 6. EFFECT OF DIGITAL FILTERING.

The processed speech segment, now back in the time domain, is stored in a new array to await output. The loop (involving a conversion to the frequency domain, the implementation of frequency assignments and subsequent call to inverse FFT) continues until all time segments have been processed.

The Smoothing Process

This procedure of taking the speech signal "a slice at a time" for processing takes advantage of the discrete nature of the stored signal. Simply recompiling the string of processed segments assumes that the envelope of each time slice does not differ drastically from what it was before processing. This has not been found to be true in practice, however. In fact, substantial noise appearing in the output of such processing may result directly from this practice. Consider, for example, the demarcation point "A" in Figure 7a, indicating where one time segment ends (array point 19) and another begins (array point 20). After processing (see Figure 7b), the relative amplitudes across that juncture are significantly disparate from one another due to the composite spectral changes made within each segment. Analog reproduction devices (especially earphones and loudspeakers) are unable to accurately transduce such a jump in amplitude. The resultant acoustic output at such a point is a transient "pop." If, for example, the time segments are each 15 msec long, then one transient would occur every 15 msec. This translates into a 67 Hz "buzz" signal which modulates the entire acoustic output, distorting its spectral content.

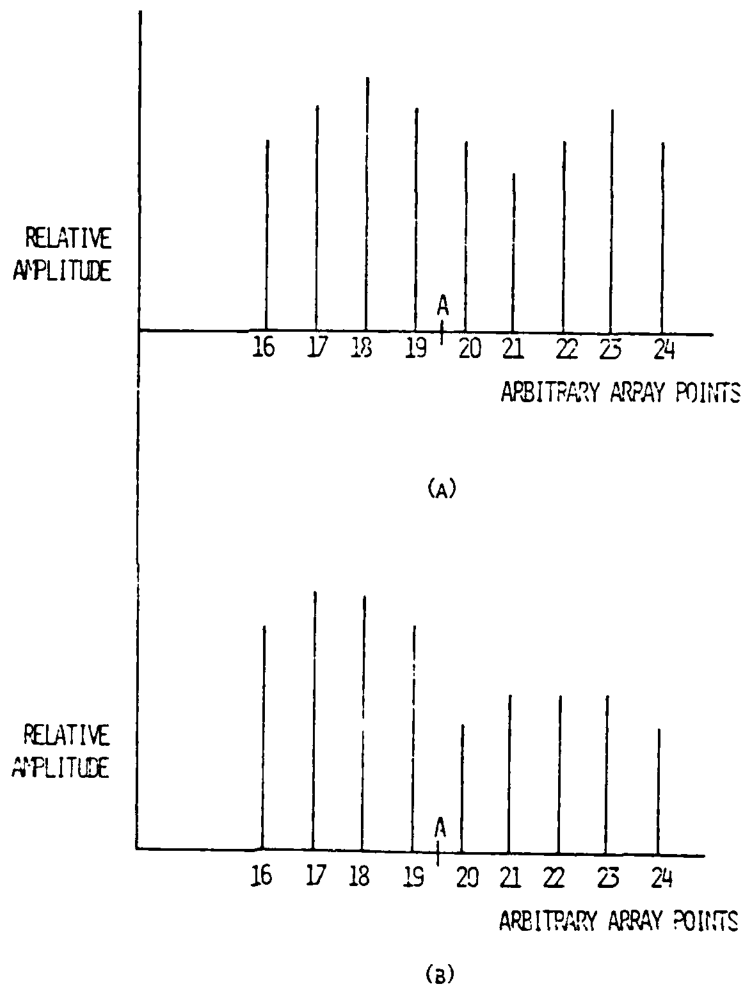


FIGURE 7. EXAMPLE OF ENVELOPE DISCONTINUITY ACROSS TIME SEGMENT BOUNDARY

To rectify this inherent situation, a software procedure was composed which will henceforth be referred to as "smoothing." The technique basically involves an isolation procedure to prevent significant envelope changes from occurring across each processed time segment boundary. The first call to FFT (henceforth known as a "pass") sends a specific number of time domain amplitude values to be converted into the frequency domain. Upon assignment of the specified spectral shape, the data are returned to the time domain via an inverse FFT call; this pass is identical to the general procedure described earlier. The second pass involves the same amount of array points in the FFT call and in the first pass, however, now the first 10 percent of the points are the same array points as the last 10 percent of the first pass' call to FFT. For example, if the first pass sends array points 1-256 to FFT, pass number 2 would send array points 230-486 to FFT. Further, pass number 3 would send array points 460-716, and so forth. This repetition between the values at the start and finish of each array isolates the boundary of each segment from large envelope discontinuities. After the last pass has been completed, the entire string of processed segments are rewritten to an output array by taking only the first 90 percent of each segment (except for the final pass, taken in its entirety). This accomplishes two goals: the repetition of small time sectors is edited, while a more nearly continuous envelope change across the boundry is approximated. The signal's amplitude vs time history is still discrete; however, these amplitudes now vary across the segment boundaries with inter-segment smoothness. Hence, the smoothing

procedure adjusts the precise features of processing-induced transients in the signal's envelope on a software level, such that auditorily perceptual "pops" are eliminated from the output while retaining the data in digital form.

Note that these manipulations (filtering assignments and smoothing procedures) all occur outside of the signal's real time. This characteristic offers several unique advantages. First, a high degree of precision is achieved during processing of the signal. Digital editing and precise spectral shaping are examples where this feature excels. Secondly, the number of different modifications greatly increases when the signal is available as discrete quanta outside of real time. Since the entire duration of the signal is accessible as a quantified whole, all dimensions of the input may be simultaneously manipulated. Finally, iterative schema may be conducted utilizing the speed of the computer's hardware to analyze different combinations of precise modifications. In many cases the experimenter does not have foreknowledge of the exact combinations needed to attain a specific output. A guessing procedure in real time is inherently limited by the need to completely process a signal each time a solution is tried. In the digital mode, however, the desired output is returned in one step since the iterations occur within the execution of the software program. Thus, the non-real time nature of digital signal processing offers greater opportunity for precise signal modification than conventional analog filters and modulators.

Chapter IV

PROCEDURES

Subjects

Twenty subjects, ten female and ten male, participated in the speech discrimination test. Ages ranged from nineteen to twenty-five years and were selected from the student population at the University. Subjects were screened for normal hearing, and those with thresholds greater than 20 dB at any frequency from 250 to 8000 Hz in octave intervals in both ears were eliminated. The right ear was the test ear for all subjects, unless the left ear showed the only normal sensitivity. All subjects were paid an hourly wage, the amount determined by the current going rate for experimental subjects at the University.

Equipment

The processed signals described in the previous chapter were generated using a hybrid computer at The Pennsylvania State University, University Park, Pennsylvania. The system is comprised of an EAI (Electronic Associates Incorporated) Model 680 analog computer interfaced with a DEC (Digital Equipment Corporation) digital computer, Model PDP-10. The A/D and D/A conversions were both performed at a rate of 20,000 samples/second. The audio output was recorded onto

magnetic tape via a Crown Model BP824 one-quarter inch tape recorder at seven-and-one-half inches per second (i.p.s.).

The discrimination tasks were performed using the apparatus diagrammed in Figure 8, including an Ampex Model AG-440B one-quarter inch tape recorder, a Maico Model MA-18 audiometer calibrated to ANSI 1969 standards, and a TDH-39 earphone fitted with an MX-41/AR cushion. The tests were performed in a Suttle Corporation Model B1 acoustically isolated quiet room.

Taped Stimulus Materials

A clinical audiometric word list was required as the input audio material to be processed. Northwestern's NU#6 word list was found to be most desirable, since it includes CCNC (consonant-consonant-nucleus-consonant) sounds as opposed to only CCVC (consonant-consonant-vowel-consonant) sounds (cf. Tillman and Carhart, 1963). In other words, it contains vowel sound combinations, i.e., nucleii such as the /ɔ/ with /i/ in the word "boil", and the /a/ with /i/ in the word "bite." These nucleii occur frequently in spoken English and a word list which includes these is especially representative of the variety of sounds naturally occurring in the language.

The tapes generated have three frequency resolutions above 500 Hz: a bandwidth equal to the resolution of the normal critical bandwidth (1X); three times the normal critical bandwidth (3X); and seven times normal (7X). The effect of the processing may be seen visually in Figures 9, 10, 11, and 12, including a comparison plot of the spectrum vs time for an unprocessed item.

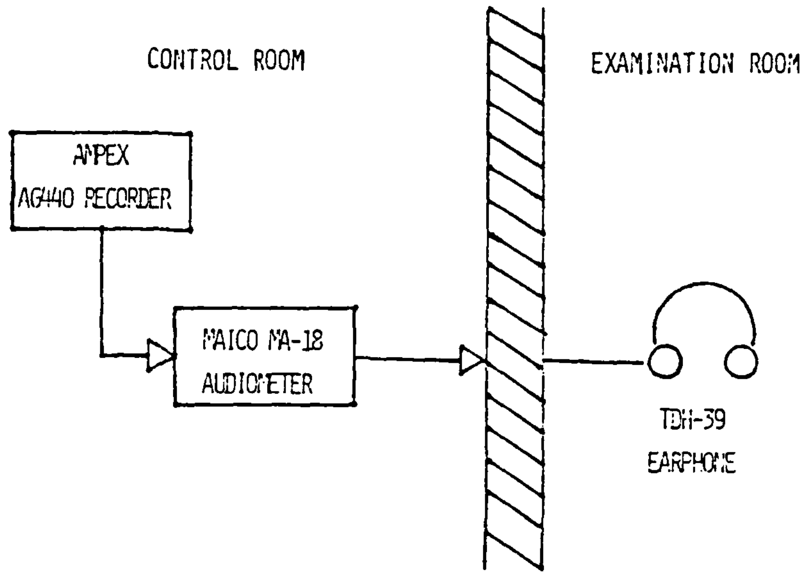


FIGURE 8. EXPERIMENTAL APPARATUS.

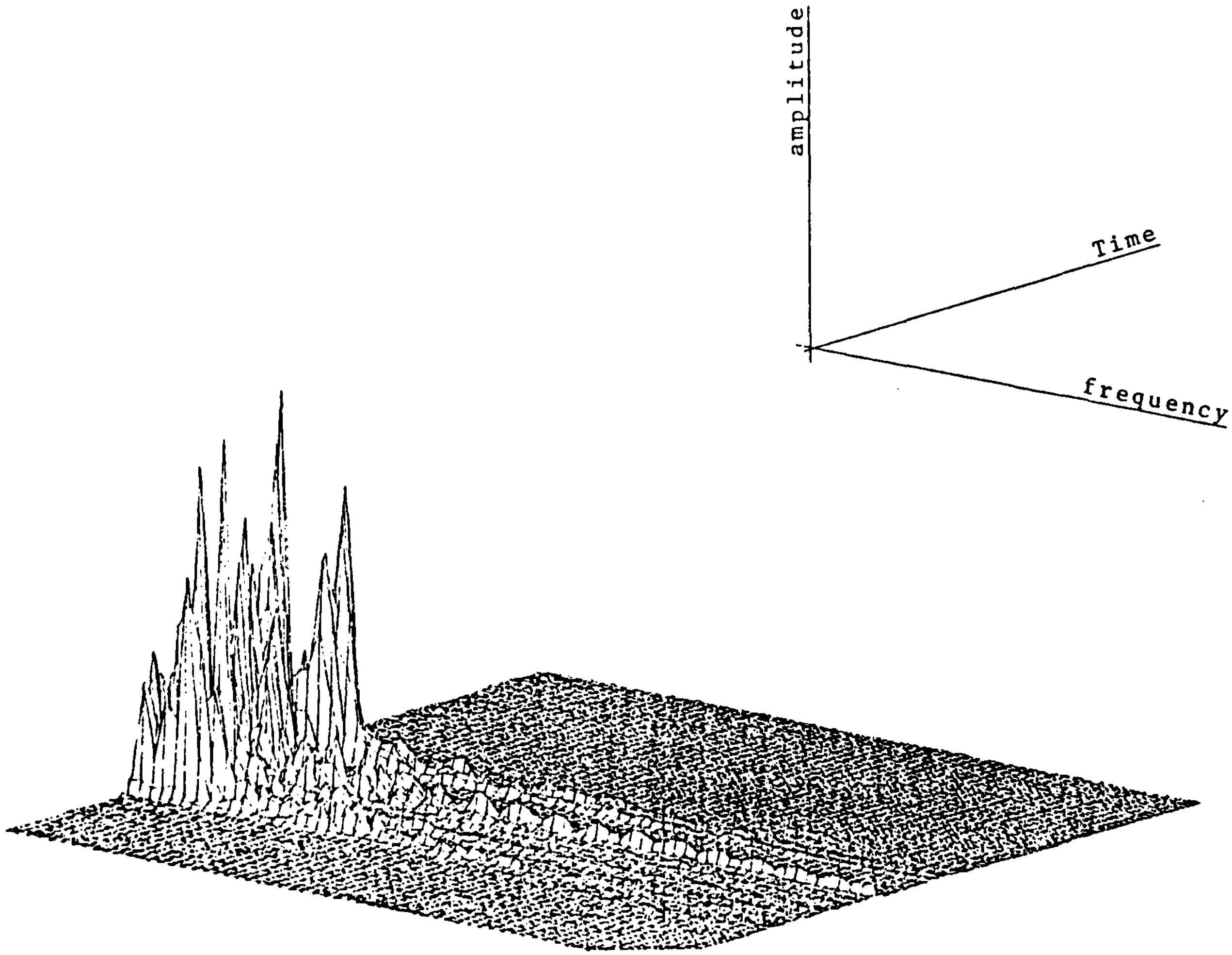


Figure 9. Three-dimensional plot of unprocessed phrase.

17-Apr-80 20:06

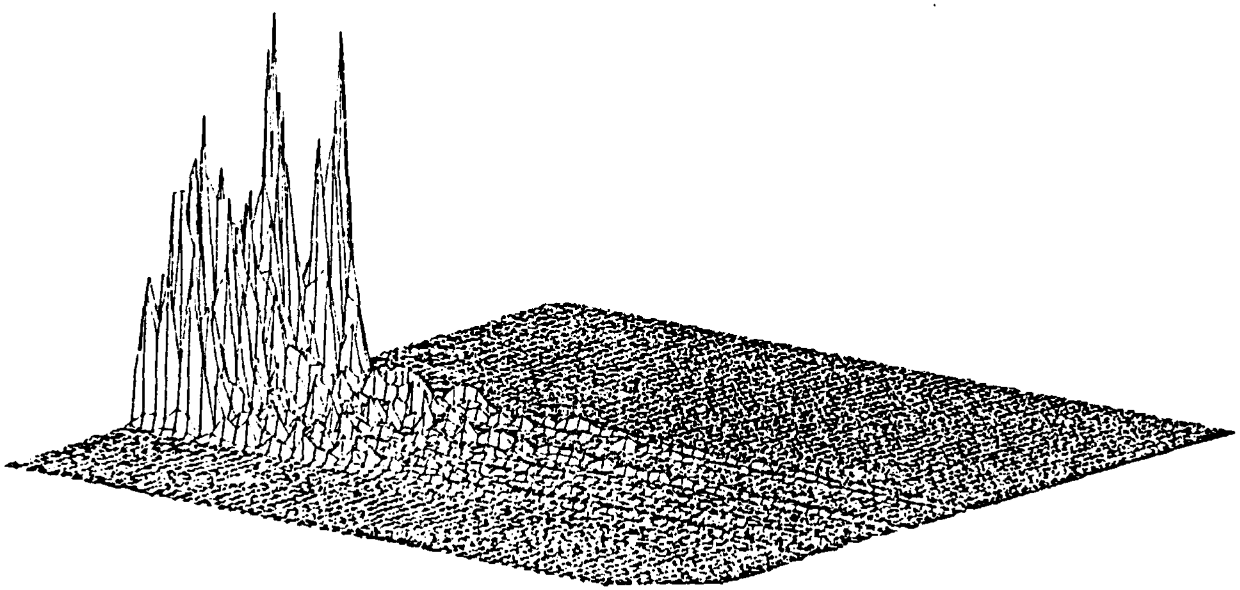


Figure 10. Three-dimensional plot of 1X condition.

17-Apr-80 20:31

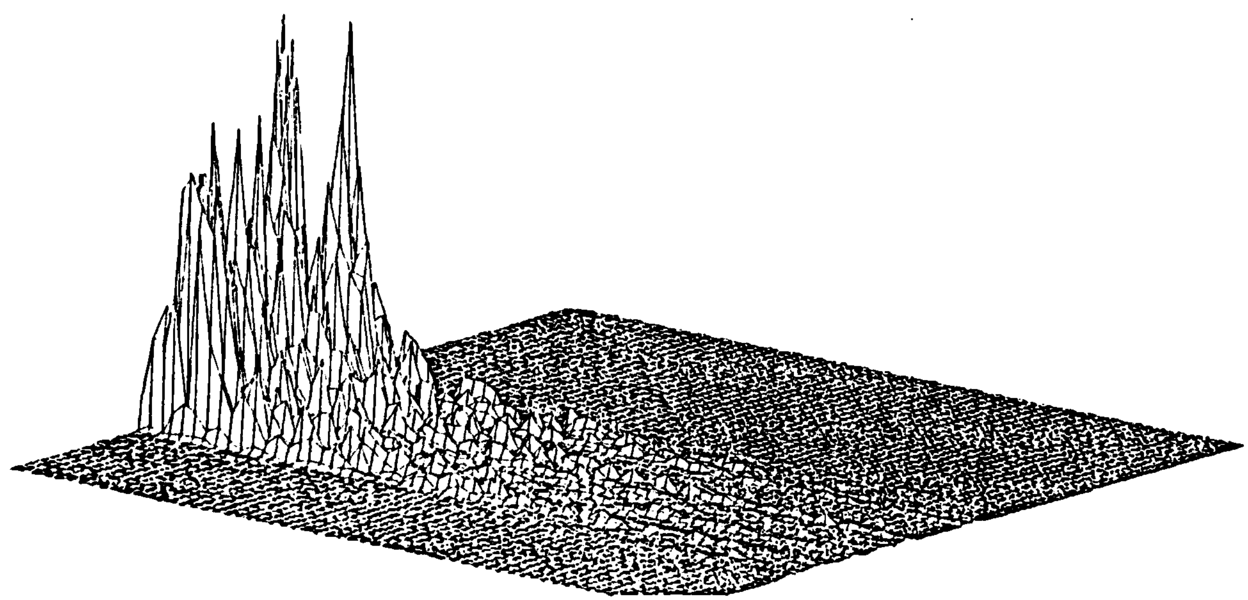


Figure 11. Three-dimensional plot of 3X condition.

17-Apr-80 20:51

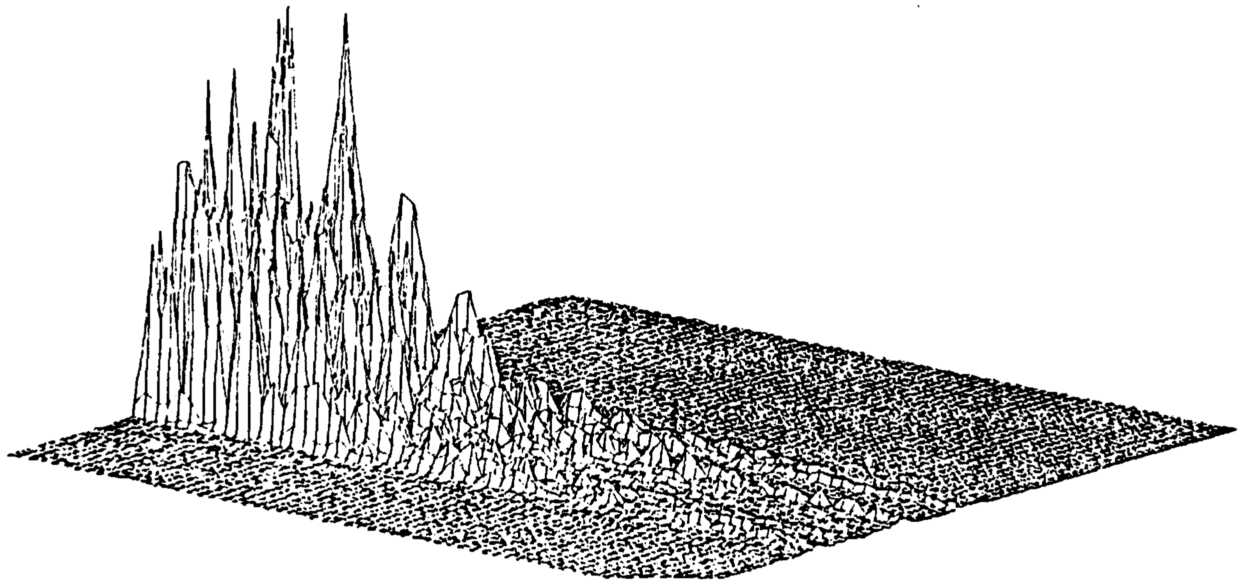


Figure 12. Three-dimensional plot of 7X condition.
17-Apr-80 21 36

Method

All subjects read and signed an informed consent document, which contained an explanation of the purpose and procedure of the study as well as an assurance of confidentiality of the data with regard to their identity. It was explained that the test basically involved listening to a standard clinical speech discrimination word list, which had been modified by a novel computer manipulation technique.

After the screening procedure (performed in the quiet room), the subjects were presented with the three fifty-word lists in the following resolution order: 7X, 3X, 1X. Since the same word list was used for each allowed resolution bandwidth, this sequence was chosen in order to minimize learning effects. The signals reached the earphone at a level of 50 dB HL and a signal-to-noise ratio of +30 dB. White noise was utilized as the masking source. Subjects were provided with three separate answer sheets and asked to write down the word they felt was said, guessing when necessary. Two points were scored for each correct identification, zero for an incorrect or blank answer. Each subject thus had three percentage scores, one for each list.

Chapter V

RESULTS

By comparing the plots of various processed phrases (see Figures 9, 10, 11, 12), the effect of the processing may be seen visually. The reader should keep in mind that in the two pathologic conditions (3X and 7X) the bands overlap each other since those in the 1X case are edge to edge. This overlap of bands induces more area to each spectral time segment (compare Figure 9 with Figure 12), since those discrete energy regions are integrated more than once. Whereas the computer interprets this effect as adding more amplitude area to a three-dimensional array, the auditory system perceives this as a loudness increase. In addition, the peaks of the phrase "Say the word, 'date'" are more noticeably rounded in the 3X and 7X filtering conditions (Figures 11 and 12) than for the unprocessed case (Figure 9), indicating the reduced resolution for the widened critical band conditions.

The group means for the discrimination tests are given in Table 2. The cell means represent the average discrimination score attained during each bandwidth-controlled presentation. An analysis of variance was performed on the discrimination data from this simple randomized design (cf. Lindquist, 1953). The results are summarized

Table 2

Mean Discrimination Scores for Several Bandwidth Resolutions

	Bandwidth Resolution		
	1X	3X	7X
Mean Discrimination Score	90.5	85.1	56.8

in Table 3. The analysis of variance indicated significant differences among the bandwidth conditions. The Tukey Multiple T-Test [a wholly significant difference test (cf. Glass and Stanley, 1970)] demonstrated a significant, monotonic, decreasing trend as bandwidth conditions were increased from 1X through 3X to 7X. That is:

a. the 3X condition scores were significantly lower than the 1X condition scores, and

b. the 7X condition scores were significantly lower than the 3X condition scores (and the 1X condition scores).

Table 3

Analysis of Variance Summary Table

Source	df	ss	ms
Treatments	2	17,876	8938*
Within Groups	57	513	9
Total	59	18,389	

*significant using F-test (cf. Lindquist, Chapter 3, 1953)

Chapter VI

DISCUSSION AND CONCLUSIONS

The peripheral auditory system performs a preliminary place-specific frequency analysis of incoming acoustic signals. An efferent feedback loop pathway carries out a selective inhibition of frequency specific afferent fibers. The limit to which frequency information may be gated is called the critical band and has been observed in such psychoacoustic contexts as masking, loudness, and musical consonance. More importantly, the critical band mechanism performs noise band limiting and harmonic discrimination, both of which are crucial for the correct perception of such complex acoustic stimuli as speech. Cochlear pathologies can affect the integrity of the critical bandwidth mechanism, which in turn can incur deficits in these functional characteristics due to widening and overlapping of the critical bands. The effect on the listener is one of resolution loss, in which their frequency resolving power is insufficient to enable discrimination of speech from an entire acoustic stimulus. To determine the unknown degree to which such a pathology has manifested itself, this preliminary research focused on a deductive approach, whereby the bandwidth resolution of a presented speech stimulus was controlled as the perceptual response was monitored.

To generate such bandwidth controlled stimuli, a digital signal processing scheme was composed, taking advantage of the precise and diverse signal modification capabilities of a discrete system. Such processing algorithms are not without usage requirements and inherent processing limitations, which are reflected in the presence of the Nyquist criterion, time segment rules, and the smoothing process. Even with these requisites, the digital approach offers greater signal processing opportunities than conventional analog filters and modulators.

The use of resolution-limited tapes on subjects with normal hearing yields preliminary data by which such a processing scheme may be evaluated. In an attempt to avoid ceiling effects, signals were presented at a signal-to-noise ratio of +30 dB. The subject performances for the 1X condition were not significantly different from 100 percent. This result indicates that the signal-to-noise condition used in this study was not rigorous enough to eliminate ceiling effects under the narrowest resolution condition.

The performance of the listeners for the 3X and 7X condition was significantly poorer than their discrimination performance at the 1X condition. Preliminary item comparison reveals that the errors of the normal listeners at the 7X condition are similar to those of sensorineural listeners with finer resolution conditions (Bienvenue et al., 1980). This trend is in conformance to the expected performance on these tests; however, this must be viewed as a preliminary conclusion since this trend is based upon the slopes given only by two pathologic resolution discrimination scores.

Further support for this conclusion stems from observation of a phenomenon occurring in the test signals as a result of processing.

The reader will recall that the amplitude area of the array under pathologic conditions (i.e., 3X and 7X cases) is increased compared to that for the normal condition (i.e., 1X case). This phenomenon was shown to result from the integration of energy more than once when that energy was located at a frequency which was common to more than one band in the widened critical band case. Whereas the computer interprets this effect as adding more amplitude area to a three-dimensional array, the auditory system perceives this as a loudness increase. This abnormal perception of a loudness increase for a given input amplitude is akin to the observed phenomenon of recruitment. Thus, there are two inherent features arising from this processing scheme that systematically model phenomenon that are known to occur in the sensorineural hearing impaired population; the processing scheme thereby warrants further study as a model of auditory processing for speech.

Further research in this area is certainly warranted and should include:

1. More rigorous signal-to-noise ratios when testing normal subjects to minimize ceiling effects;

2. The use of sensorineural hearing-impaired subjects whose bandwidth distortion is determined beforehand by other psychophysical techniques and compared to the processing condition at which they first exhibit a detriment in performance;

3. Inclusion of more bandwidth conditions (2X and 5X, for example) to yield a more accurate determination of performance trends.

Summary

This research has been a true interdisciplinary endeavor, drawing from such schools of thought as engineering, physiology, physics, mathematics, and psychology. In conclusion, this thesis has demonstrated:

1. the feasibility of generating speech signals of varying bandwidth resolutions equal to and greater than the normal critical bandwidth using a novel digital signal processing algorithm, specifically,

    a. bandwidth resolution limited signals such as those described in Chapter III have been produced;

    b. the processing scheme allows for variation of bandwidths from the normal critical band to integer multiples of the critical band;

    c. the stored array of processed speech signals demonstrates an abnormal growth of amplitude in the pathological cases (3X and 7X); a phenomenon akin to the recruitment seen in sensorineural hearing loss;

2. that normal listeners hearing the 3X and 7X lists demonstrate decremental performance in speech intelligibility, specifically,

    a. speech intelligibility scores for a pathologic processing scheme equivalent to a pathologic listening condition (3X or 7X) are comparable to scores obtained by a sensorineural hearing impaired listener presented with unprocessed, equivalent word lists;

    b. as the bandwidth resolution allowed to the normal listeners was systematically widened, the intelligibility scores showed a monotonic decreasing trend;

3. that the proposed processing scheme is a reasonable approximation to the modeling of hearing for the purposes of speech intelligibility in both the normal and pathologic cases, and that this approach warrants further study and development.

REFERENCES

Bennett, T., and others. "Procedures for Characterizing Certain Effects of Prolonged Noise Exposure." Journal of the Acoustical Society of America, 63, Supplement 1, 1978.

Bienvenue, G., and P. Michael. "The Temporary Effects of Short Term Noise Exposure on Masking Phenomenon." Unpublished EAL Research Project, 1977.

________ and others. "Effects of Asymptotic Noise Exposure on a Battery of Tests." Paper presented at the meeting of the American Audiology Society, December 12, 1977.

________ and others. "Digital Processing of Speech Materials in the Study of Sensorineural Hearing Impairment." Journal of the Acoustical Society of America, 67, Supplement 1, 1980.

Bonding, P. "Critical Bandwidth in Presbycusis." Scandinavian Audiology, 8, 1979.

Capps, M. and H. Ades. "Auditory Frequency Discrimination after Transection of the Olivo-Cochlear Bundle in Squirrel Monkeys." Experimental Neurology, 21, 1968.

Cooley, J., P. Lewis and P. Welch. "The Finite Fourier Transform." IEEE Transaction on Audio and Electroacoustics, 17(2), 1969.

________ and J. Tukey. "An Algorithm for the Machine Calculation of Complex Fourier Series." Mathematical Computations, 19, 1965.

Curtis, J., Cleft Palate and Communication, eds. D. Sherman and D. Spriestersbach, (New York: Academic Press, 1968).

Dallos, P., The Auditory Periphery, (New York: Academic Press, 1973).

Desmedt, J., and P. Monaco. "Mode of Action of the Efferent Olivo-Cochlear Bundle in the Inner Ear." Nature, 192, 1961.

Dewson, J. III. "Efferent Olivo-Cochlear Bundle: Some Relationships to Stimulus Discrimination in Noise." Journal of Neurophysiology, 31, 1968.

Fernandez, C. "The Innervation of the Cochlea." Laryngoscope, 61, 1951.

Fex, J. "Auditory Activity in Centrifugal and Centripetal Cochlear Fibers in Cats." ACTA Physiologica Scandinavia, 55, 1962.

________. "Efferent Inhibition in the Cochlea Related to Hair Cell Deactivity: A Study of Postsynaptic Activity of the Crossed Olivo-Cochlea Fibers in the Cat." Journal of the Acoustical Society of America, 41, 1967.

Fletcher, H. "Auditory Patterns." Review of Medical Physiology, 12, 1940.

Fowler, E. "Marked Deafened Areas in Normal Ears." Transactions of the American Otological Society, 18, 1928.

________. "Measuring the Sensation of Loudness." Archives of Otolaryngology, 26, 1937.

French, N., and J. Steinberg. "Factors Governing the Intelligibility of Speech Sounds." Journal of the Acoustical Society of America, 19, 1947.

Glass, G. and J. Stanley., Statistical Methods in Education and Psychology, (Englewood Cliffs: Prentice-Hall, 1970).

Goldstein, M. "The Auditory Periphery." Medical Physiology, ed. V. Mountcastle, (Saint Louis: Mosby, 1968).

Greenwood, D. "Auditory Masking and the Critical Band." Journal of the Acoustical Society of America, 33, 1961.

Haggard, M. "Feasibility of Rapid Critical Bandwidth Measurements." Journal of the Acoustical Society of America, 55, 1974.

Ladefoged, P., Elements of Acoustic Phonetics, (Chicago: Chicago University Press, 1962).

Lindquist, E., Design and Analysis of Experiments in Psychology and Education, (Boston: Houghton Mifflin, 1953).

Lorente de No, R. "The Sensory Endings in the Cochlea." Laryngoscope, 47, 1937.

Michael, P., and G. Bienvenue. "A Procedure for the Early Detection of Noise Susceptible Individuals." American Industrial Hygiene Association Journal, 1976.

Milner, P., Physiological Psychology, (New York: Holt, Rinehart and Winston, 1970).

Minifie, F., T. Hixon, and F. Williams., Normal Aspects of Speech, Hearing and Language, (Englewood Cliffs: Prentice-Hall, 1973).

Moore, B., Introduction to the Psychology of Hearing, (Baltimore: University Park Press, 1977).

Morton, J. and A. Carpenter. "Experiments Relating to the Perception of Formants." Journal of the Acoustical Society of America, 35, 1963.

Pickles, O., and S. Comis. "Role of Centrifugal Pathways to Cochlear Nucleus in Detection of Signals in Noise." Journal of Neurophysiology, 36, 1973.

Plomp, R. "The Ear as a Frequency Analyzer." Journal of the Acoustical Society of America, 36, 1964.

________ and W. Levelt. "Tonal Consonance and Critical Bandwidth." Journal of the Acoustical Society of America, 38, 1965.

Rabiner, L., and R. Schafer., Digital Processing of Speech Signals, ed. A. Oppenheim, (Englewood Cliffs: Prentice-Hall, 1978).

Ramon Y Cajal, S., Beitrag zur Studium der Medulla Oblongata, (Leipzig: Barth, 1896).

Rasmussen, G. "The Olivary Peduncle and Other Fiber Projections of the Superior Olivary Complex," Journal of Comparative Neurology, 84, 1946.

Remez, R. "Adaptation of the Category Boundary Between Speech and Nonspeech: A Case Against Feature Detectors." Haskins Laboratories: Status Report on Speech Research SR-50, 1977.

Scharf, B. "Critical Bands." Foundations of Modern Auditory Theory, ed. J. Tobias, (New York: Academic Press, 1970).

Singleton, R. "An Algorithm for Computing the Mixed Radix Fast Fourier Transform." IEEE Transactions on Audio and Electroacoustics, 17(2), 1969.

Spoendlin, H. "The Innervation of the Cochlea Receptor." Basic Mechanisms in Hearing, ed. A. Moller, (New York: Academic Press, 1973).

Tillman, T., and R. Carhart. "An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words (Northwestern University Auditory Test No. 6)." Technical Report, SAM-TR-66-55, United States Air Force School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas, 1966.

Trahiotis, C., and D. Elliott. "Behavioral Investigations of Some Possible Effects of Sectioning the Crossed Olivo-Cochlear Bundle." Journal of the Acoustical Society of America, 48, 1970.

Woolsey, C., and E. Walzl. "Topical Projection of Nerve Fibers from Local Regions of the Cochlea to the Cerebral Cortex of the Cat." Bulletin of Johns Hopkins Hospital, 1942.

Zwicker, E. "Die Verdecklung von Schmalbandgerauschen durch Sinustone." Acustica, 4, 1954.

Zwicker, E. "Ueber Psychologische und Methodische Grundlagen der Lautheit." Acustica, 8, 1958.

________ and R. Feldtkeller. "Ueber die Läutstarke von Gleichförmigen Geräuschen." Acustica, 5, 1955.

DISTRIBUTION

Commander (NSEA 09G32) Naval Sea Systems Command Department of the Navy Washington, DC 20362	Copies 1 and 2
Commander (NSEA 0342) Naval Sea Systems Command Department of the Navy Washington, DC 20362	Copies 3 and 4
Defense Technical Information Center 5010 Duke Street Cameron Station Alexandria, VA 22314	Copies 5 through 16

DATE
FILMED

– 8